REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
		R&D Status Report 1/1/97-3/31/97

4. TITLE AND SUBTITLE
Applications of the Theory of Distributed and Real-Time Systems to the Development of Large-Scale Timing Based Systems

5. FUNDING NUMBERS
F19628-95-C-0118

6. AUTHOR(S)
Nancy Lynch

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Department of the Airforce
Electronic Systems Center (AFMC)
Hanscom Air Force Base, MA 01731

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
N/A

12a. DISTRIBUTION/AVAILABILITY STATEMENT
No limits of disclosure.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

14. SUBJECT TERMS	15. NUMBER OF PAGES
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	III

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements.***

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	PR	- Project
G	- Grant	TA	- Task
PE	- Program Element	WU	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD	- See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE	- See authorities.
NASA	- See Handbook NHB 2200.2.
NTIS	- Leave blank.

Block 12b. Distribution Code.

DOD	- Leave blank.
DOE	- Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
NASA	- Leave blank. .
NTIS	- Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

April 15, 1997

Mr. Harry Koch
ESC/ENS
5 Eglin Street, Building 1704
Hanscom Airforce Base, MA 01731-2116

Dear Mr. Koch:

This letter contains our R & D Status Report covering the period from January 1, 1997 to March 31, 1997 for Contract F19628-95-C-0118, entitled "Applications of the Theory of Distributed and Real-Time Systems to the Development of Large-Scale Timing-Based Systems".

Technical Progress

In the following report, more information about the people mentioned can be found on our group's "people" page, at URL http://theory.lcs.mit.edu/tds/people.html.

I. Modelling and verification tools

- Garland and Lynch have completed a tentative design of a programming language for I/O automata, which they call "IOA". This language allows simple abstract descriptions for distributed systems and is intended for use in system development, testing, and verification.

  During this quarter, Vaziri tested the formal notations provided by IOA by using them to transcribe most of the algorithm descriptions in Lynch's book, *Distributed Algorithms*. As a result of this exercise, Garland made several changes in the grammar for IOA and updated the parser. He also added static semantic checks for the portions of the language used to specify abstract data types.

  A graduate student, Xioawei Yang, began work on the static semantic checks related to I/O automata. When Yang decided to join another research group, Garland finished what she had begun, and then turned the implementation of further semantic checks over to Svetoslav Tzvetkov, an undergraduate research assistant.

  Vaziri and Garland continued writing a user's manual for IOA. The manual is almost complete. Petrov and Vaziri worked on a translation scheme from IOA to the input language of the model checker SPIN.

  A new master's student, Anna Chefter, will join the project in June and begin developing a simulator for I/O automata.

- Jensen continued work on techniques for integrating model checking and theorem proving methods for verification of concurrent systems. He has proved property-preserving abstraction theorems for the I/O automata framework and has used his theorems in the verification of a concurrent read/write algorithm. A paper entitled "Abstraction Methods for Model Checking in the I/O Automata Framework", documenting the above work, is in progress.

- Segala and Lynch finally finished rewriting the paper "Liveness in Timed and Untimed Systems" (also co-authored with Gawlick and Sogaard-Andersen) for journal submission. This paper presents a compositional treatment of liveness properties for both untimed and timed distributed systems. During this reporting period, some of the key definitions involving "receptiveness" of components of timed systems were significantly simplified.

II. Algorithms and impossibility results

- Hoest and Shavit continued their work on a mathematical complexity theory for fault-tolerant asynchronous systems. They have used topological models and methods to analyze time complexity in the *iterated immediate snapshot* model, a restricted type of atomic snapshot shared memory model. They obtained tight bounds for the approximate agreement problem, and a fundamental time vs. number of names tradeoff for the renaming problem. A paper detailing these results, entitled "Towards a Topological Characterization of Asynchronous Complexity" was submitted to PODC'97. Hoest and Shavit are currently working on extending their complexity theory to other types of shared memory models.

- Della Libera and Shavit's work on reactive diffracting trees was accepted to SPAA'97. The paper describes a new version of the diffracting tree synchronization primitive that grows and shrinks according to the load on the data structure. They are now writing a full journal verson of the paper.

- Shavit and Zemach completed work on a highly concurrent priority queue design based on their earlier "combining forests" data structure. They completed empirical evaluations of the design using the Proteus simulator, and are writing a technical report for conference submission. The next step will be design modifications and empirical tests on the MIT Alweife machine here at MIT. Shavit and Zemach also completed the "Diffracting Trees" manuscript, which will appear in ACM TOCS. Their paper with Upfal of IBM Almaden on the journal version of their SPAA 96 paper providing a mathematical model for analyzing diffracting tree performance was accepted to a special issue of the journal Mathematical Systems Theory.

Finally, they submitted to PODC'97 a paper on a new "wait-free" sorting algorithm – that is, one that will take logarithmic parallel time and will run (though slightly less effectively) even if many processes fail.

- Touitou and Shvartsman completed a suite of simulations validating the earlier theoretical results of Lynch, Shavit, Shvartsman and Touitiou showing that many important classes of the highly concurrent data structures used for counting and load balancing exhibit nearly linearizable behavior. A journal paper is being prepared for submission (a preliminary report appeared in PODC'97).

- Shvartsman completed the manuscript, *Fault-Tolerant Parallel Computation.* This monograph synthesizes the latest results for parallel computation in the presence of failures, restarts and delays. The manuscript is currently in the production phase at Kluwer Academic Publishers.

- Chlebus, De Prisco and Shvartsman developed a new fault-tolerant algorithm for the *Do-All* problem of performing n tasks using p message-passing processors under the constraint of maintaining message and work efficiency. This is the first algorithm for the problem that efficiently deals with processor restarts. A manuscript is in preparation.

III. Applications

A. Distributed system building blocks

- Shvartsman and Oleg Cheiner, an M.Eng. student, continued experimentation using a prototype distributed algorithm based on the eventually serializable data service of Fekete, Gupta, Luchangco, Lynch and Shvartsman, presented in PODC 96. The fully-distributed algorithm implements several optimizations. It runs on a LAN of Unix workstations and uses the MPI message passing system. Recent development includes the implemention of an additional synchronization mechanism that allows greater concurrency. It trades off replica synchronization for performance. Emprical study is in progress.

- The paper by Lynch and Shvartsman "Robust Emulation of Shared Memory Using Dynamic Quorum-Acknowledged Broadcasts" was accepted by FTCS'97. The paper defines a new reconfigurable quorum-based broadcast-convergecast communication primitive. The primitive is used to obtain new fault-tolerant distributed implementations of serializable shared read/write memory. The paper is being revised to incorporate referee comments.

- Fekete, Lynch and Shvartsman completed a conference version of their paper on group communication services and submitted it to PODC'97. This paper contains automaton-based specifications for group communication primitives such as those used in the Isis, Transis, Horus and Psynch systems. In particular, the paper includes specifications for a virtually synchronous group communication (VSGC) service and for a totally ordered broadcast service. Fekete et al. have modelled an algorithm, derived from one of Dolev and his students, that uses VSGC to implement totally ordered broadcast. They have a good outline of an assertional correctness proof for this algorithm, plus specifications and proofs giving performance and fault-tolerance properties.

  During this reporting period, they edited and polished the paper. They also began work with Myla Archer at the Naval Research Lab, on verification of the safety proofs using PVS. Talks about this work at several institutions and a visit from Ken Birman have helped to publicize the work.

- Khazan continued his work on modeling a load-balancing replicated database server that relies on an underlying VSGC layer to achieve efficiency and fault-tolerance. So far he has developed an automaton-based specification and implementation for this problem, as well as a function which he believes to be a "forward simulation" from the latter to the former. Currently, he is working on a proof of the forward-simulation result which will imply correctness of the algorithm. The next steps will be a performance and fault-tolerance analysis of the implementation.

- De Prisco continued work on finishing his M.S. thesis, which contain a model, proof and analysis for Lamport's Paxos algorithm for fault-tolerant distributed consensus. He submitted a paper based on this work (co-authored with Lampson and Lynch) to PODC '97.

B. Multiprocessor shared memory models

- Luchangco continued his work on developing a theory of precedence-based memory models, which generalize multiple processor memory models, and abstract away system implementation details. During this reporting period, he defined a generalized notion of sequential consistency, and a weak consistency requirement called per-location sequential consistency, and established sufficient conditions under which the two types of memory are indistinguishable to clients. He also proved that an algorithm used by the Cilk system implements a per-location sequentially consistent memory.

- Frigo and Luchangco have begun to develop a theory of "computation-centric memory models", which characterize memories from the programmer's point of view. A computation is a generalization of an instruction stream. Memory models are expressed in terms of these computations, allowing the programmer to reason about what a program specifies rather than about low-level system details. They have defined sequential consistency in this framework, along with several weak consistency models, and have proved some properties of these models, as well as relationships among them.

B. Automated Transportation Systems

- Livadas continued his work on the use of Hybrid I/O Automata to model and prove correctness of vehicle protection subsystems, as used in the Raytheon Personal Rapid Transit project. His model allows composition of protectors that depend on each other's correctness. The correctness proofs of the various protectors are facilitated by the proof of correctness of an abstract protector — a generic protector that captures the abstract functionality of a protector without considering the particular physical plant and protector details. Correctness proofs for protectors preventing overspeed and collisions both for a straight track and a general track topology involving multiple Y-shaped merges and diverges have been completed. Some technicalities involving the abstract protector and protector composition remain to be addressed.

- Dolginova and Lynch continued their work on modeling and analyzing safety criteria for the platoon maneuvers for the California PATH intelligent highway project, using Hybrid I/O Automata. Prior to this reporting period, the ideal case, and some more complicated cases involving delays and sensor uncertainty, were modeled and verified. This quarter, they began considering other kinds of uncertainty, such as the uncertainty in break performance. The paper on Safety Verification of Automated Platoon Maneuvers was presented in the HART'97 workshop in Grenoble France in March '97. It has also been partly rewritten for a technical report.

- Lygeros initiated two new projects involving automated transportation systems – the first involving Automated Highway Systems (AHS) and the second Air Traffic Management Systems (ATMS). This work is intended not just to contribute results about the safety of these systems, but also to help establish important theoretical links between computer science and control techniques for designing and analyzing hybrid (discrete/continuous) systems. In particular, the plan is to combine techniques from theoretical computer science (such as invariant assertions and simulation relations) that work well for dealing with discrete dynamics, with

techniques from control theory (such as optimal control and game theory) that are powerful for continuous dynamics. Specifically:

- **AHS research**: Lygeros has begun studying the problem of emergency deceleration of a string of vehicles. This is a hybrid problem because of the interaction between the deceleration process (continuous) and the vehicle collision dynamics (discrete). He is trying to establish conditions under which such a maneuver can be executed safely, i.e., in such a way that any collisions that may occur are at low relative velocities. In this reporting period, he developed a model for describing the evolution of this system and investigated necessary and sufficient conditions for safety for a particular deceleration maneuver (all vehicles braking as hard as possible). This work is especially important for AHS architectures that involve platooning of vehicles.

- **ATMS research**: Lygeros has begun considering the problem of verifying the TCAS conflict detection/resolution algorithms. This is a hybrid problem because of the interaction between the aircraft dynamics (continuous) and the inter-aircraft communication protocols (discrete). The goal is trying to verify that the newest proposed conflict resolution algorithm guarantees safety, i.e., that under reasonable assumptions, it maintains a minimum separation between the aircraft. In this reporting period, Lygeros developed a preliminary model for the system. He is currently working on extracting the communication protocols from the latest TCAS documents and fitting them into the model. This work is important to the area of ATMS because it provides ways of formally proving the correctness of the protocols before they are deployed. (Currently the protocols are only tested in simulation, a process that does not provide absolute guarantees.)

C. Communication

- Smith continued to work on his PhD. thesis entitled "Formal Verification of TCP and T/TCP". The thesis contains a formal verification of TCP with unbounded counters and TCP with bounded counters. It also shows that the T/TCP protocols does not guarantee at most once semantics, but satisfies a weaker specification.

- Smith and Lynch continued to refine their impossibility result for the "at-most-once fast delivery problem". This problem is the one the TCP/IP transport level protocol T/TCP is designed to solve. The impossibility result states that if the client and server do not have accurate clocks, then no protocol can solve this problem. There is a paper in progress, entitled "The Impossibility of at-most-once fast message delivery."

D. Probabilistic Systems

- Segala and Lynch continued working on finishing up their paper on modelling, verifying, and analyzing the Aspnes-Herlihy randomized consensus protocol. Besides a proof of this particular algorithm, many useful techniques for analyzing randomized distributed systems have been produced in the course of this work. A paper was submitted to PODC '97.

Special Programs and Major Items of Equipment

None.

Changes in Key Personnel

None.

Trips, Talks and Conferences

1. Nancy Lynch. "Specifying and Using a Virtually Synchronous Group Communication Service." Florida International University, January 1997. Invited Lecturer.
2. Nancy Lynch. "Specifying and Using a Virtually Synchronous Group Communication Service." Yale University, February 1997. Invited Lecturer.
3. Nancy Lynch. "Specifying and Using a Virtually Synchronous Group Communication Service." ARPA Networking PI Meeting, Baltimore, Maryland, March 1997.
4. Nancy Lynch. "Mathematical Modelling/Specification/Verification/Performance Analysis/Fault-Tolerance Analysis for Network Services." ARPA Active Nets Workshop, March 1997.
5. Nancy Lynch. "Specifying and Using a Virtually Synchronous Group Communication Service." VERIMAG School on Methods and Tools for Verification of Infinite State Systems, Grenoble, France, March 1997. Guest Speaker.
6. Alex Shvartsman. "Efficient and Fault-Tolerant Parallel Computation." Information Technology Institute of the University of Valencia, Spain, January, 1997.
7. Alex Shvartsman. "Distributed Systems and Building Blocks." Information Technology Institute of the University of Valencia, Spain, January, 1997.
8. Alex Shvartsman attended ICDT'97 and CP'97, Delphi, Greece, in January, 1997.
9. Alex Shvartsman. "Efficient Parallel Computation with Processor Failures and Delays." Heinz Nixdorf Institut, Universität-GH Paderborn, in January, 1997. Shvartsman was a visitor at the Institute during the second half of January.

10. John Lygeros. "Multi-objective Hybrid Controller Synthesis". Presented at *Conference on Hybrid and Real Time Systems (HART97)* in Grenoble, France, March 1997.

11. Ekaterina Dolginova. "Safety Verification for Automated Platoon Maneuvers: A Case Study." Presented at *International Workshop on Hybrid and Real-Time System (HART97)*, Grenoble, France, March 1997.

12. Nir Shavit. "Towards a Topological Characterization of Asynchronous Complexity. " Presented at MIT Theory of Computation Seminar series, Cambridge, MA, March 1997.

Areas of Concern

None.

Statement of Sufficiency

The contractually prescribed effort appears to be sufficient to achieve the objectives of this contract.

Degrees awarded

None.

Related Accomplishments

During this reporting period the following papers have been submitted for publication, accepted for publication, or published:

[1] Alan Fekete, Nancy Lynch, and Alex Shvartsman. Specifying and Using a Partitionable Group Communication Service. Submitted for publication, January 1997.

[2] Roberto DePrisco, Butler Lampson, and Nancy Lynch. Revisiting the Paxos Algorithm. Submitted for publication, January 1997.

[3] N. Shavit, E. Upfal, and A. Zemach. A Wait-Free Sorting Algorithm. Submitted for publication.

[4] Gunnar Hoest and Nir Shavit. Towards a Topological Characterization of Asynchronous Complexity. Submitted for publication.

[5] Anna Pogosyants, Roberto Segala, and Nancy Lynch. Verification of the Randomized Consensus Algorithm of Aspnes and Herlihy: A Case Study. Submitted for publication.

[6] Alan Fekete, M. Frans Kaashoek, and Nancy Lynch. Implementing Sequentially Consistent Shared Objects Using Broadcast and Point-to-Point Communication. Submitted for journal publication.

[7] Giovanni Della Libera and Nir Shavit. Reactive Diffracting Trees. *Proceedings of the 9th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA).* To appear.

[8] N. Shavit and E. Upfal, and A. Zemach. A Steady State Analysis of Diffracting trees. *Mathematical Systems Theory.* Special Issue. To appear.

[9] Nancy Lynch and Alex Shvartsman. Robust Emulation of Shared Memory Using Dynamic Quorum-acknowledged broadcasts, *Twenty-Seventh Annual International Symposium on Fault-Tolerant Computing* (FTCS'97), Seattle, Washington, USA, June 1997. To appear.

[10] Ekaterina Dolginova and Nancy Lynch. Safety Verification for Automated Platoon Maneuvers: A Case Study. *International Workshop on Hybrid and Real-Time System (HART97)*, Grenoble, France, March 1997.

[11] Nir Shavit and Dan Touitou. Software Transactional Memory. *Distributed Computing.* 10:2, January/February 1997. Special Issue.

Papers in progress

Oleg Cheiner. "Implementation and Evaluation of an Eventually-Serializable Data Service." Masters thesis.

B. Chlebus, R. De Prisco, and A. Shvartsman. "Work in a Message-passing Environment Prone to Processor Failures and Restarts."

Roberto De Prisco. "Revisiting the Paxos algorithm." Masters thesis.

Matteo Frigo and Victor Luchangco. "Computation-Centric Memory Models."

Stephen J. Garland, Nancy A. Lynch, and Mandana Vaziri, "IOA: a Formal Language for I/O Automata."

Gunnar Hoest. "Towards a Topological Characterization of Complexity in Asynchronous, Distributed Systems." Masters thesis.

Henrik Jensen. "Abstraction Methods for Model Checking in the I/O Automata Framework."

Roger Khazan. "Group Communication as a Base for a Load-Balancing, Replicated Data Service." Masters thesis.

Jon Kleinberg, Hagit Attiya, and Nancy Lynch. "Trade-offs between Message Delivery and Quiesce Times in Connection Management Protocols." Journal version.

Carolos Livadas. "Verification of Automated Vehicle Protection Systems." Masters thesis.

Victor Luchangco. "Precedence-Based Memory Models."

Victor Luchangco. "Building Blocks for Distributed Computing Applications." PhD thesis.

John Lygeros and Nancy Lynch. "Conditions for Safe Platoon Deceleration."

John Lygeros and Nancy Lynch. "Formal Verification of the TCAS Conflict Resolution Algorithms." An extended abstract was submitted to an invited session for the 1997 Conference on Decision and Control.

Nancy Lynch, Roberto Segala, Frits Vaandrager, and H. B. Weinberg. "Hybrid I/O Automata." Journal version.

Nancy Lynch, Nir Shavit, Alex Shvartsman, and Dan Touitou. "Timing Conditions for Linearizability in Counting Networks." Journal version.

Nancy Lynch and Sergio Rajsbaum. "On the Borowsky-Gafni Simulation Algorithm." Journal version.

Mark Smith. "Formal Verification of TCP and T/TCP." PhD thesis.

Mark Smith and Nancy Lynch. "The Impossibility of At-Most-Once Fast Message Delivery."

Awards:

- In March 1997, Dr Lygeros was awarded the Eliahu Jury Award by the Department of Electrical Engineering and Computer Sciences of the University of California, Berkeley, for "outstanding research in the area of Systems, Communications, Control or Signal Processing".

Sincerely,

Nancy Lynch
NEC Professor of Software Science and Engineering

Electrical Engineering and Computer Science
(617)253-7225
lynch@theory.lcs.mit.edu

MIT Laboratory for Computer Science

Applications of the Theory of Distributed Real-Time Systems
To the Development of Large-Scale Timing-Based Systems

Prof. Nancy Lynch, Principal Investigator

R & D Status Report
Program Financial Status
ARPA Contract # F19628-95-C-0118
CLIN # 0002
Quarterly Report (1/97 - 3/97)

	Planned Expenditures	Actual Expenditures at Report Date	% Completion	Budget At Completion	Latest Revised Estimate	Remarks
Total Base Contract	858,443	268,661	31.30%	858,443	858,443	
Current Funding Profile	574,447	268,661	46.77%		268,661	*
Equipment	35,308	0	0.00%			**

* Data reflects all received funding. Current funding is sufficient through 10/97.

** Equipment funding is for 3 budgeted workstations. To date, none have been purchased.